ISSN 0281-9864

Remarks on the Status of Inference in the Area of Knowledge Representation

Christopher Habel

KOGNITIONSVETENSKAPLIG FORSKNING

Cognitive Science Research

Lund University
Sweden

Remarks on the Status of Inference
in the Area of
Knowledge Representation

Christopher Habel

1987 No. 20

Communications should be sent to:
Bernhard Bierschenk
Department of Psychology
Paradisgatan 5
Lund University
S-22350 Lund, Sweden

Abstract

The concept of inference is one of the global concepts used for the explanation of cognitive processes. There exist mainly two types of characterization: the logical and the psychological. These different characterizations are based on the difference between inference and rule of inference.
Information processing systems can be formalized as inferential systems, i.e. systems with inferential processes. The fundamental concepts of this formalization, those of dynamical inferential systems and time-restricted derivations, both based on inferential processes are described in detail.

Remarks on the status of Inference in the area of Knowledge Representation

1. Preliminaries on *Cognitive Science*

One of the fundamentals of Cognitive Science (CS) - and thus of Cognitive Psychology and Artificial Intelligence (AI) - is the view of minds as information processing systems (IPS), cp. Haugeland (1978). The strongest, i.e. most far-reaching, version of *Cognitivism* can be characterized by Newell's (1980)

Physical Symbol System Hypothesis

"... humans are instances of physical symbol systems, and by virtue of this, mind enters into the physical universe." (p. 136)

Thus the subject in Cognitive Science is the inquiry of information processes from different points of view and with different goals. Considering the whole spectrum from Cognitive Psychology to applied Artificial Intelligence this leads to such different goals as the explanation of processes in the human mind, on one hand, and the use of information processes in applied AI-systems, on the other hand. (The current state of the art is dominated by a strategy of investigation, which Churchland (1984; p. 106) characterizes - w.r.t. AI - as the "piecemeal approach": only very specific processes are investigated; a global view of the whole phenomenon of human intelligence is out of the range of our scientific knowledge yet.) Analogously we also have a spectrum of experiments to lay an empirial base for the investigation of mind. This reaches from experiments with natural subjects (human or animals) in the psychological tradition to computer systems in the simulation mode. (From now on I will neglect the application-oriented parts of AI and therefore the application mode of AI-systems too.) Later on in the present paper I will sketch a third type of evidences, which could be called *theoretical evidences*. Such evidences are only possible by virtue of strict formalisms. And at this point, two older relatives of the twins Cognitive Psychology and Artificial Intelligence, namely Mathematics and Logics, will enter the stage.

Following the information processing paradigm of the human mind it is necessary to postulate system-internal models of real or fictional worlds. These internal models are built up by (formal) *knowledge entities* on a semantic or intentional level. (On this level cp. Pylyshyn 1984; p. 210-211). Beyond these representations, procedures (*procedure* is used here in a non-technical sense) are needed to work on the representations. By this the subject area in *Representation of Knowledge* can be outlined: the investigation of these representations and the operations on them to change the internal models.

2. Inferences

In the present section I will give emphasis to the most relevant (from my point of view) concept in the area of knowledge representation: *inference*. The concept of *inference* is one of the global concepts (see above) used for the explanation of general cognitive processes. The notion of *inference* can be found all over Cognitive Science but without a unique and well-established definition. There exist mainly two types of characterization:

the **logical characterization**:
"THE RULES OF INFERENCE...... amount to directions as to how sentences already known as true may be transformed so as to yield new true sentences." (Tarski, 1965; p. 47)

the **psychological characterization**:
"The process of a conclusion, or a conclusion reached, on the basis of previously made or accepted judgements." (Drever, 1964; p. 136.)

These roots lead to a converged AI characterization (or one of *formal Cognitive Science*):

"Inferences are well-defined changes of attitudes to knowledge entities"

At this point one important distinction has to be called to notice, namely that between the concepts *rule of inference* and *inference*. The former refers to an instruction or advice how to move from a set of attitudes to some knowledge entities to another set of attitudes to the same or other knowledge attitudes. This means, that rules of inference can be seen as *inference-tickets*, which license the change of attitudes. (The notion of *inference-ticket* is due to G.Ryle (1949 p. 117).) The latter concept (of *inference*) refers to the process or action of transforming attitudes itself.

Taking this distinction into consideration is very important for Cognitive Science. *Rules of inference* are knowledge entities of a specific type, i.e. they are part of the system's (human's) knowledge base. In contrast to this, *inferences* are performed by the system during information processing. The former have to do with the system's potential competence or capacities the latter with it's actual performance.

At this point a relevant pair of questions arises:

Do humans use valid, i.e. formally justified, inference rules?
Are human's inferences valid?

If the answer would be "Yes!" all would be very nice, that means, no problems would appear. But a lot of experiments show that humans are bad in doing valid inferences; cp. Johnson-Laird's (1983) chapter "How to

reason syllogistically". Since the answer has to be "No!", in Cognitive Science we have two possible reactions with respect to this fact, which I want to name as 'conservative' and 'liberal':

conservative reaction:
Formal systems should not reproduce human's mistakes.

liberal reaction:
Formal systems should describe and explain the inference systems of humans, although they are not valid from a logical point of view.

Because the conservative opinion is - implicitly or explicitly - prevalent in the disciplines of logic and formal semantics nearly no cognitively relevant contribution to the topic of inference processes have come from these fields. On the other hand, most cognitive psychologists, cp. Johnson-Laird (1983), agree with the liberal view, but formal inference-systems describing and explaining the human competence and performance are rare up to now. To develop such systems is one main topic of Cognitive Science in the future. With respect to language processing this insight is the base of the important remarks on *Semantic Intuition* by the logicians and semanticists Barwise and Cooper (1981):

> "While it is seldom made explicit, it is sometimes assumed that there is some system of axioms and rules of logic engraved on stone tablets - that on inference in natural language is valid only if it can be formalized by means of these axioms and rules. In actuality, the situation is quite the reverse. The native speaker's judgement as to whether a certain inference is correct, whether the truth of the hypothesis implies the truth of the conclusion, is the primary evidence of a semantic theory ..." (Barwise/Cooper 1981; p. 201-202)

To sum up the situation: In Cognitive Science and its related disciplines different concepts exist of an *inferential system* dependent on the main research topics of the discipline in question. Needed is a unified treatment of inferences and reasoning processes (cp. Section 4 of this paper).

3. The basic structure of inferential systems

The mostly investigated and best understood inferential systems are the calculi of formal logic. I will use the structure of these systems to exemplify and summarize the basic ideas and to introduce the notation, which is used later on.
Following traditional logics, eg. Carnap (1939) or Tarski (1965), a *calculus* with respect to a formal language **L** is defined by

$\mathbf{A}$, a set of sentences, the *axioms*
and $\mathbf{R}$, a set of rules (of inference).

Based on this, the concepts of *derivations* and *proofs* are defined in the wellknown way. The set of all sentences which are derivable from **A** by means of rules from **R** is named as *set of theorems*:

$$\mathrm{Th}(\mathbf{A}, \mathbf{R}) := \mathrm{Th}_{\mathbf{R}}(\mathbf{A})$$

The natural extension of this idea to applied formal systems, i.e. *formal theories*, leads to the *deductive method* ; cp. Carnap (1939), Tarski (1965), Kalish/Montague (1964). The language **L** has to be enriched by nonlogical constants leading to a language **L'**. Axioms and inference rules are defined over **L'**, especially nonlogical axioms are used. By this deductive method wide ranges of mathematics and the sciences can be treated in a strictly formal manner.
Let us now continue with a *psychological interpretation* - with respect to IPSs - of formal systems.

A concerns the factual knowledge of the IPS which the system contains in an explicit way (see below),

R contains the IPS's rules of inference, which determine the system's inferential capacity,

$\mathrm{Th}_{\mathbf{R}}(\mathbf{A})$ refers to the implicit knowledge of the system, i.e. the knowledge entities which can be reached by derivations (or inferences).

Before I will go on with the explicit - implicit dichotomy I will enter into a sketchy discussion of the question what can be inferred (and by which methods inferences are performed) by formal systems based on standard logics.
As first test case the natural quantifier MOST is to be mentioned. On the one hand, humans perform MOST-inferences very well. Therefore there is the need of formal inference-systems to describe and explain the phenomena related with MOST-expressions. On the other hand, there exist no adequate, fully satisfactory treatment of MOST in any logical approach (cp. Rescher (1962), Kaplan (1966) and Barwise/Cooper (1981) on negative results of MOST-formalizations.) Cushing's (1987) two-predicate-scope MOST-quantifier is - though interesting from a logial point of view - limited w.r.t. explanations of human's MOST-inferences.

The second example for beyond-traditional-logic inferences concerns SEEING. As Barwise/Perry (1983) demonstrate a *logic of seeing* does exist. They postulate an inventory of principles for the behaviour of perceptual reports. As an example I will only mention their

Principle of Veridicality:

If **b** sees φ, then φ. (Barwise/Perry 1983; p. 181, 187)

(Remark: This case concerns so called '*NI-Perceptual Reports*', where 'NI' stands for 'naked infinitive'. In the present paper I will not discuss advantages and disadvantages of Barwise and Perry's proposal, which is under a controversial discussion in linguistics; cp. Higginbotham (1983). For my argumentation the type of formalization, not the specific solution, is used.)

The principle of veridicality can be transformed into an inference rule, e.g.

$$\mathrm{SEE}\,(b,\varphi) \rightarrow \varphi$$

Using the same type of formalization an analogous rule should be postulated for *knowing*, namely

$$\mathrm{KNOW}\,(a,\varphi) \rightarrow \varphi$$

which is a notational variant of Hintikka's (1962) condition (C.K.) (p. 43) in a logic of knowledge.

What is the moral of these examples of KNOWING and SEEING? As I described in more detail in another paper (Habel, 1983) there exists a tendency for some operators (or concepts) to change their status from beeing *non-logical* to becoming *logical*. The development of epistemic logic is possible only by treating **k** and **b** (for KNOWING and BELIEVING) as logical constants. The trend in future formal systems will be to use more and more today-non-logical concepts, e.g. SEEING, as logical ones. This means that new logics, e.g. a logic of perceptual reports has to be developed in a formal manner. (The treatment of SEEING in situation semantics by Barwise/Perry - see above - is just a first step on this way.)

I now will come back to the explicit-implicit distinction. Using the interpretation given at the beginning of the present section, namely,

$\mathbf{A}$	~ axioms	~ explicit knowledge
$\mathrm{Th}(\mathbf{A})$	~ theorems	~ implicit knowledge

the relation between these types of knowledge, symbolized by

$$\mathbf{A} \vdash \mathbf{A}^* := \mathrm{Th}_{\mathbf{R}}(\mathbf{A})$$

is the relation of *derivability*. In contrast to this, from a psychological (or cognitive) point of view another relation and a third type of knowledge has to be emphasized, namely the *actually derived knowledge* $\mathbf{A}\#$ produced by and connected via the relation of *derivation*. This situation (i.e. relation) can be symbolized by

$$\mathbf{A} \mathrel{|\!\sim} \mathbf{A}\#$$

analogously to the logical relation mentioned above.

At this point of argumentation some remarks about my use of the logical notions is necessary: The "one-step" relation between two sets of formulas F_1, F_2 which is induced by application of one inference rule with respect to the former set F_1 resulting in F_2 is usually named as deduction or derivation and symbolized by "$\vdash$". Often an attribute 'direct' or 'in one step' is used here. From a logical point of view sequences of direct derivations are as interesting as one-step derivations. Therefore the notion of *derivation* (in one or more steps) is introduced, formally by means of the transitive closure of the relation of direct derivation. (Beside of derivation also *proof* is used.) But actual derivations or proofs are not the major topics of interest for a logician; instead of the actual derivation the possibility to derive or prove is the topic of investigation. Therefore with respect to *proof* the notion of *provability* is interesting. Because of this specific focussing on the possibility of the existence of a derivation, I give emphasis to *derivability* if I describe the logical way of investigating derivations and inferences.

At this point it should be mentioned that the question whether the relation of derivability has to be established in a constructive way is intensively discussed in the area of logic and proof theory. The details of this discussion, which belong to the foundation of mathematics, are beyond the subject of the present paper.

Questions of the type

How many steps of derivation are needed to ...?

Which way of derivation is used ...?

(Both with respect to a pair consisting of a set of axioms and a sentence) are very seldom investigated in traditional logic. (A very interesting note on this topic is Boolos' (1987) paper, which describes a first-order inference rule not feasible practically but usable in a second-order variant.)

4. On the dynamics of inferential systems

Given an inferential system $\mathbf{I} = \langle \mathbf{A}, \mathbf{R} \rangle$ we can investigate inferential processes with respect to **I**. Such investigations are concerned with the questions stated at the end of the preceding section. Looking at a natural inferential system, e.g. a human, the dynamical properties of the system are relevant and therefore topics in Cognitive Science. Furthermore, we have to take into consideration the distinction (mentioned above) between

$\mathbf{A}^{\#}$ ~ the actually derived knowledge

and $\mathbf{A}^{*}$ ~ the potentially derivable knowledge

This distinction is interesting only from the computational point of view: $\mathbf{A}^{*}$ is the topic of formal logic whereas $\mathbf{A}^{\#}$ is of major interest in CS and AI. A distinction analogous to that between $\mathbf{A}^{*}$ and $\mathbf{A}^{*}$ is the main topic of Levesque's (1984) "Logic of implicit and explicit belief". The major differences between his and my approach are:

Levesque distinguishes explicit and implicit beliefs without dealing with the question of processes which make implicit beliefs explicit.

Levesque does not deal with processes which change beliefs, i.e. he does not consider the dynamical properties of knowledge sets (see below).

Only with respect to both types of implicit knowledge, namely $\mathbf{A}^{\#}$ and $\mathbf{A}^{*}$, questions of the type

Which knowledge entities are derived from $\mathbf{A}$ at a specific point of time?

are sensible.

Beyond the dynamics with respect to specific inferences, i.e. sequences of actions, from a cognitive point of view there exists a second type of dynamics in inferential systems, namely with respect to the systems. To clarify the two types of dynamics I introduce a further theory-internal entity, a *set of points of time*

$$T = \{t_i\}$$

I do not want to develop a "theory of time" here. Instead of this I list some relevant properties of T

i. T is an ordered set (ordering $\leq$)
ii. T is infinite
iii. T is discrete

These properties of T reflect the use of *time* in the following: points of time are seen as indices of states of the inferential system. (More elaborated *time-theories* are developed by and described in Rescher/Urquhart (1971) and van Benthem (1983).) Furthermore, I claim the existence of a minimal element t_0, which corresponds to initial state. Thus it is possible to use the natural numbers, as a canonical index set; cp. also Habel (1985). An interval $T' \subset T$ of time points, i.e. sets

$$T' := [t_i, t_j] = \{t / t_i \leq t \;\&\; t \leq t_j\}$$

will be named as *time-span*. t_i and t_j are called beginning time, **b(T')**, and ending time, **e(T')**, respectively.

By means of **T** there exists a natural way to speak about the "inference rule r_i used at t_i during an inference process", and to restrict inference processes with respect to time resources. Given an inferential system $\langle \mathbf{A}, \mathbf{R} \rangle$ and a time-span $T' \subset \mathbf{T}$ it is now sensible to define

$$\mathbf{Th}_{\mathbf{R}}(\mathbf{A}, T') := \{S / \mathbf{A} \vdash_{T'} S\}$$

similar to the usual definition of **Th** based on the relation $\vdash$ (derivability) with <u>one</u> important distinction, namely that sets of *time-restricted theorems* are based on *time-restricted derivations* characterized as follows:

$S_1 \vdash_{T'} S_2$ iff

There exists a sequence of inference rules which constitutes $S_1 \vdash S_2$ and there is an index-mapping from T' into this sequence.

Remark: "time-restricted set of theorems" and "time-restricted derivation" are here not defined in a strict sense; the "definitions" above have to be seen as characterizations. In other words, in the present paper only the idea of a definition is given in form of sketchy remarks. From these characterizations it should be clear, that the cognitively relevant set **A#** representing derived knowledge is of the type described above, i.e. **A#** contains the knowledge entities derived during some time-span T'. Note, that though **A#** represents derived knowledge with respect to a specific time-span, nothing is said here about the reasons for deriving just the knowledge entities of **A#** - **A** (the difference of the knowledge sets) during T'. Questions of this type are not topic of the present paper.

Let me return to **T** again. Up to now, we used only one sub-set, namely the time-span **T'**, of **T**. In the next extension of the theoretical inventory I will make use of specific coverings of **T**:

Let $\mathcal{T} = \langle T_1, \ldots, T_k \rangle$ be a sequence of time-spans with
$\mathbf{e}(T_i) = \mathbf{b}(T_{i+1})$ for i=1,...k-1 and
$\mathbf{e}(T_1) = t_0$, the minimal time-point.
$\mathcal{T}$ is called an initial covering of **T**.

By means of such initial coverings it is possible to characterize recursively a sequence of inferential systems and sets of knowledge entities:

Given $\langle \mathbf{A}, \mathbf{R} \rangle$, an inferential system,
and $\mathcal{T} = \langle T_1, \ldots, T_k \rangle$, an initial covering

$$\mathbf{A}_1 := \mathbf{A}$$
$$\mathbf{A}_{i+1} := \mathrm{Th}_{\mathbf{R}}(\mathbf{A}_i, T_i)$$

The dynamical development of derived knowledge can be exemplified graphically as follows

$$\begin{array}{llllllll}
\mathbf{A}_1 & \vdash_{T1} & \mathbf{A}_1^{\#} & & & & & \\
 & & =: & & & & & \\
 & & \mathbf{A}_2 & \vdash_{T2} & \mathbf{A}\#_2 & & & \\
 & & & & =: & & & \\
 & & & & \mathbf{A}_3 & \vdash_{T3} & \mathbf{A}\#_3 & \\
 & & & & & & =: & \\
 & & & & & & \ldots \quad \ldots & \mathbf{A}^*
\end{array}$$

Thus we have changed from the investigation of one inferential system (or formal theory) to sequences of inferential systems. From a cognitive point of view it is obvious that the situation in natural inference systems are much more complicated than the sketch (given above) demonstrates. The most important further extension would concern the dynamics of $\mathbf{R}$. This means, in contrast to the case described above in which every $\langle \mathbf{A}_i, \mathbf{R} \rangle$ uses the same set of inferential rules $\mathbf{R}$, we have to consider the development (over time) of $\mathbf{R}_i$ too; i.e. interesting (because cognitive realistic) inferential systems have a behaviour of the $\langle \mathbf{A}_i, \mathbf{R}_i \rangle$-type. Steps from $\mathbf{R}_i$ to $\mathbf{R}_{i+1}$ can be seen as induced by learning processes. (Cp. Emde/Habel/Rollinger 1983 , Habel i. prep.)

5. Conclusion: Mutual relation between logics and psychology

In this concluding section I will summarize the main results and insights which should be influential between the poles 'logic' and 'psychology' of the Cognitive Science spectrum. Furthermore, I will formulate some mutual requirements.
Logic and 'theoretical AI' have a lot of results on the formal properties of information processing systems, especially inferential systems. These results concern e.g. problems of decidability, generative capacity and complexity. From a cognitive point of view there are some desiderata, e.g. with respect to the dynamics of inferential processes and systems. Furthermore, traditionally only some phenomena of human knowledge processing are subject of logic. For example, beyond the standard quantifiers 'all' and 'some' only few results exist (cp. the remarks on MOST in section 3). In contrast, psychology offers empirical data and theoretical models with respect to derivations by natural inference systems. The formal analysis of the psychologist's models requires

methods from logic and theoretical AI. That formal analysis is useful demonstrates the example of 'most'-inferences. Up to now no fullly adequate formal treatment of 'most' does exist. The most promising approaches can be classified into two types of *cardinality approaches* vs. *default-approaches*. The former class contains the formalizations of Rescher (1962), Kaplan (1964), Barwise/Cooper (1981), the latter Reiter (1980). Both types of formalizations have been investigated with respect to formal properties, e.g. decidability or existence of proof-procedures. Such formal results concern the topic of *theoretical evidences* mentioned in section 1: e.g. lower "theoretical cost" of one formal system w.r.t. another could possibly correspond to lower cognitive costs of the natural inference systems in question. And by this could be explained why the cheaper system is selected in spite of missing validity (in a formal sense). (The usefulness of theoretical evidences for Cognitive Science will be the topic of a future paper.)

References:

Barwise, John/Cooper, Robin (1981):
Generalized Quantifiers and Natural Language. Linguistics and Philosophy 4 (2) . 159-219.

Barwise, John/Perry, J. (1983):
Situations and Attitudes. Bradford Books: MIT Cambridge, Mass.

van Benthem, J.F. (1983):
The Logic of Time. Reidel: Dordrecht

Boolos, George (1987):
A Curious Inference. Journal of Philological Logic 16. 1-12.

Carnap, Rudolf (1939):
Foundation of Logic and Mathematics. Encyclopedia of Unified Science, Vol I, No 3: Chicago

Churchland, Paul M. (1984):
Matter and Consciousness. MIT-Press: Cambridge, Mass.

Cushing, Steven (1987):
Some Quantifiers Require Two-Predicate Scopes . Artificial Intelligence 32, 259-67

Drever, James (rev. ed. 1964):
The Penguin Dictionary of Psychology. Penguin: Harmondsworth.

Emde, Werner/ Habel, Christopher/ Rollinger, Claus-Rainer (1983):
The discovery of the equator or concept driven learning. Proc. 8th IJCAI. 455-458.

Habel, Christopher (1983):
Logische Systeme und Repräsentationsprobleme
in: B. Neumann (ed): GWAI-83. Springer: Berlin. 118-42

Habel, Christopher (1985):
Referential Nets as Knowledge Structures. in: Th. Ballmer (ed): Linguistic Dynamics. de Gruyter: Berlin. p. 62-84

Habel, Christopher (i. prep.):
A note on the dynamics of rule systems.

Haugeland, John (1978):
The Nature and Plausibility of Cognitivism. in: J. Haugeland (ed): Mind Design. Bradford: Montgomary Vt. 243-81

Higginbotham, James (1983):
The logic of perceptual reports: An extensional alternative to Situation Semantics. Journal of Philosophy 80(2). 100-27

Hintikka, Jaakko (1962):
Knowledge and Belief. Cornell Univ. Press: Ithaka, NY.

Johnson-Laird, P.N. (1983):
Mental Models. Cambridge UP: Cambridge.

Kalish, Donald /Montague, Richard (1964):
Logic-Techniques of formal reasoning. Harcourt, Brace & World: New York.

Kaplan, David (1966):
Rescher's plurality quantification. Journal of Symbolic Logic 31. 153-4

Levesque, Hector J. (1984):
A Logic of Implicit and Explicit Belief. AAAI 1984: 198-202

Newell, Allen (1980):
Physical Symbol Systems. Cognitive Science 4: 135-83

Pylyshyn, Zenon (1984):
Computation and Cognition. Bradford MIT-Press: Cambridge, Mass.

Reiter, R. (1980):
A Logic for Default Reasoning. Artificial Intelligence 13: 81-132.

Rescher, Nicholas (1962):
Plurality Quantification. Journal of Symbolic Logic 27: p. 373-4.

Rescher, Nicholas/Urquhart, Alasdair (1971):
Temporal Logic. Springer: Wien.

Ryle, Gilbert (1949/1973):
The Concept of Mind. Penguin University Books: Harmondsworth, Middlesex.

Tarski, Alfred (1965):
Introduction to Logic. Oxford UP: New York

Author's Note:

The author's address is: Christopher Habel, Universität Hamburg, Fachbereich Informatik, Bodenstedtstr. 16, D-2000 Hamburg 50, Fed. Rep. Germany.

An earlier version of this paper was presented at a symposium on Knowledge Representation at the 35th Annual Meeting of the German Psychological Society (Deutsche Gesellschaft für Psychologie) in September 1986.